# NEW SOLUBLE AND STABILIZED TRIMERIC FORM OF GP41 POLYPEPTIDES

The instant invention is directed to soluble and stabilized trimeric forms of the envelope glycoprotein gp41 of HIV-1 and to their use as vaccine agent.

Human immunodeficiency virus type 1 (HIV-1) encodes a 160 kDa envelope glycoprotein (gp160) precursor, which is proteolytically cleaved into the exterior (gp120) and transmembrane (gp41) glycoproteins.

In the glycoprotein mature envelope, the gp120 glycoprotein remains associated with the gp41 ectodomain through a noncovalent interaction. The native HIV-1 envelope glycoproteins exist as trimers that consist of three gp120 and three gp41 subunits and is anchored in the viral or infected cell membrane by the gp41 transmembrane region.

It has been shown that the binding of gp120 to the CD4 receptor induces conformational changes that promote subsequent interaction with one of a number of chemokine receptors (CXCR4, CCR5...). These binding events trigger conformational changes in gp41. In particular, studies by X-ray crystallography and nuclear magnetic resonance indicate that the viral envelope glycoprotein gp41 exists in at least three conformations, a native conformation (spike), a prefusogenic metastable conformation which is converted to a thermostable fusogenic "three hairpin" conformation following a triggering event, such as binding of HIV virus particle to the membrane of target cells.

So, the binding of gp120 to cellular coreceptors induces the gp41 conversion of a prefusogic form to a fusogenic form.

The linear organization of the gp41 includes a fusion peptide, an ectodomain (a N-terminal coiled-coil, a disulfide-bonded loop region, and a C-terminal α-helical segment) and a transmembrane domain.

In the fusogenic six-helix bundle, three N-terminal helices form a trimeric coiled-coil, and three C-terminal helices pack in the reverse direction into three hydrophobic grooves on the surface of the coiled-coil. This helical-hairpin structure corresponds to the fusion-active conformation of gp41. Because the membrane anchor and the fusion peptide of the gp41 ectodomain are embedded in the viral and target cell membranes, respectively, the formation of the fusogenic hairpin structure results in the colocalization of the two membranes and thus overcomes the energy barrier for membrane fusion.

The envelope glycoproteins represent the only realistic viral target for vaccine-induced neutralizing antibody responses because they promote viral membrane fusion through receptor-mediated conformational change and they are expressed on the surface of both virions and infected cells. Monomeric HIV-1 gp120 and derivatives were initially considered to be principal vaccine candidates. However, HIV-1 gp120 is highly variable and has repeatedly proven to be an immunogen ineffective at eliciting neutralizing antibodies against clinical HIV-1 isolates. Few of the antibodies raised by gp120 monomers effectively bind assembled HIV-1 envelope glycoprotein trimers.

In contrast, gp41 is an extremely immunogenic glycoprotein, inducing antibodies in essentially all HIV-infected individuals.

The ectodomain of gp41 is the most conserved region in HIV-1 envelope, which otherwise exhibits considerable genetic diversity even among closely related isolates.

Furthermore, the gp41 performs a critical role in maintaining the conformation and infectivity of the HIV virion.

The antibodies targeting the six-helix bundle (fusogenic form) and prehairpin (prefusogenic form) structures arrest fusion under certain conditions. Antibodies having access to prehairpin and six-helix bundles conformations of gp41 would be capable of inhibiting gp41-mediated fusion. Furthermore, the six-helix bundle is an extremely stable structure.

Those observations allow considering the gp41 six-helix, under a modified form or not, as an attractive target for drugs and vaccine development.

In US 6 455 265, the inventors showed that some gp41 derivatives could be particularly efficient for obtaining vaccine for preventing the pathogenic effects related to a retroviral infection with the proviso that the corresponding polypeptides have epitopes having a modified antigenicity so as to obtain a differential immune response with respect to the viral envelope, and some self-proteins.

More precisely, they discovered that conserved and immunodominant regions of the retroviral envelope could be responsible for harmful autoimmune phenomena, particularly in the case of the gp41 retroviral envelope. The inventors have observed that certain immunodominant regions of the gp41 exhibit three-dimensional structural

analogies and/or cross-reactivity with certain regions of a protein of the human immune system, and in particular IL-2.

Accordingly, they proposed modified polypeptides obtained by modifying the antigenicity of the concerned epitope of the envelope protein, in order to obtain a differential immune response with respect to the viral envelope protein and these proteins of the human immune system.

Generally, the gp41 can be produced in baculovirus or mammalian cells but the yield is lower than in *E.coli*. Furthermore, the glycosylation in baculovirus or mammalian cells is different from the glycosylation of human cells and is not necessary for the immunogenicity of the protein. Gp41 is in fact very immunogenic without glycosylation.

However, recombinant HIV ectodomain of gp41 produced in *Escherichia coli* forms insoluble precipitates (aggregates of gp41 trimeric form) at neutral pH.

The instant invention is more precisely directed to propose stabilized hydrosoluble forms of gp41 protein, in particular of gp41 protein derivatives and more particularly of derivatives as disclosed in US 6 455 265.

Unexpectedly, the inventors have discovered that it was possible to decrease significantly the hydrophobicity of the loop, resulting in an improvement of the solubilisation of the recombinant HIV ectodomain of gp41, without altering its immunogenic reactivity.

Accordingly, within one aspect of the invention there is provided a modified polypeptide containing at least an immunodominant region and the connecting loop between N- and C-helices of gp41 ectodomain of HIV-1, wherein the connecting loop includes at least a linker fragment having:

- a size convenient for keeping the native conformation of the interaction between N- and C-helices, and
- an hydrophily sufficient to provide a soluble and stable trimeric form to said modified polypeptide.

In the meaning of the instant invention, the expression "keeping the native conformation of the interaction between N- and C-helices" is understood to mean that the interaction between the N- and C- helices of the gp41-modified polypeptide is functionally similar or equivalent to the interaction of those helices of the wild-type gp41 polypeptide.

Within one embodiment of the invention, the linker fragment is included in the loop in substitution of all or only in part of a deleted wildtype oligopeptide.

Within another embodiment of the invention, the deleted wildtype oligopeptide is located in the region from 598 to 622, in particular in the region from 603 to 615 of the gp41 protein, according to the numbering of SEQ ID NO 1 (Fig. 2A).

Within another embodiment of the invention, the deleted oligopeptide is located in the region from 525 to 549, in particular in the region from 530 to 542, according to the numbering of SEQ ID NO 14 (Fig. 3).

Within a further embodiment, the deleted wild type oligopeptide consists of a sequence of at least 10, in particular 13 and more particularly 25 amino acid residues.

Within another embodiment, the linker fragment is an oligopeptide linker mainly based on hydrophilic amino acids residues.

Within still another embodiment of the invention, the oligopeptide linker consists of the sequence SGGRGGS as set forth in SEQ ID NO 2.

Within a second aspect of the invention, there is provided a polynucleotide encoding a modified polypeptide as disclosed above.

Within a third aspect of the invention, there is provided an expression vector comprising the following operably linked elements: a transcript promoter, a DNA segment encoding a modified polypeptide as disclosed above and a transcript terminator.

Within a fourth aspect of the invention, there is provided a vaccine for preventing the pathogenic effect related to a retroviral infection including as active material at least one modified polypeptide as disclosed above.

The definition given above implies that the polypeptide used comprises at least part of an immunodominant region of the gp41 viral envelope protein of HIV-1.

The modified polypeptide in accordance with the present invention may be, for example, the whole envelope protein of VIH-1, modified as indicated hereafter. The modified polypeptide may also be part of the envelope protein, modified as indicated hereafter, said part comprising at least one immunodominant fragment as defined hereafter. The modified polypeptide may also be a chimeric protein comprising at least part of the envelope protein, said part of the envelope protein being as defined above.

The peptide sequence of region 545-684 (SEQ ID NO 1) reproduced in the appended in Figure 2A is a consensus sequence (32 HIV-1 strain SWISS PROT). .

The sequence represented on Figure 3, SEQ ID NO 14, is derived from sequence of the gp41 of the HxB2 strain HIV virus (SWISS PROT CODE: ENV HV 1 BR), wherein the amino acid residues cysteine in positions 530 and 536 have been replaced by the amino acid residues serine.Immunodominant region refers to a peptide sequence which induces, in a great majority of cases (for example in at least 7 cases out of 10 approximately), a humoral and/or cellular response of the immune system directed against said region after immunization with a protein containing said sequence or with a peptide essentially consisting of said sequence.

In the present application, when reference is made to an immune response, without any other specific information, it is an immune response of a vertebrate, following immunization *in vivo*.

The invention makes reference to the target cells of a virus which are the cells into which the virus is capable of penetrating. The target cells of retroviruses are generally known. Viruses have the property of binding to the cells which they are capable of infecting. It is therefore optionally possible to test for, using routine experiments *in vitro*, the target cells of a virus studied.

The invention also makes reference to the cells of the host having a membrane receptor for a protein of the host. The cells of the host which have a receptor for a protein of said host are often known and, in the opposite case, it is possible, using routine experiments, to determine if a given protein binds to a certain type or cell. It is possible, for example, to use a radiolabelled protein and to determine if it binds to said cell type. It is also possible to test if the protein binds to a given membrane receptor using a cell line transfected with a gene expressing said membrane receptor.

The proteins of the host for which certain cells of the host possess a membrane receptor are mainly proteins belonging to the range of soluble protein mediators. This range includes proteins called, depending on the cases, hormones, growth factors or cytokines, although there is no distinct boundary between these various categories of mediators; see for example CAVAILLON J. M., Les Cytokines (Masson, Paris, 1966) Chapter 1, pages 1-3 and preface.

The linker fragment considered according to the invention have an overall hydrophilic character and is non or weakly immunogenic and flexible.

It is a synthetic linker and more particularly, it will be an oligopeptide linker.

As used herein, a "flexible" linker is one that lacks a substantially stable higher-order conformation in solution. Areas of local charge are to be avoided. In general, small, polar, and hydrophilic residues are preferred, and bulky and hydrophobic residues are undesirable. If the linker polypeptide includes charged residues, they will ordinarily be positioned so as to provide a net neutral charge within a small region of the polypeptide. It is therefore preferred to place a charged residue adjacent to a residue of opposite charge.

In general, preferred residues for inclusion within the linker polypeptide include Gly, Ser, Ala, Thr, Asn, and Gln; more preferred residues include Gly, Ser, Ala, and Thr; and the most preferred residues are Gly and Ser. In general, Phe, Tyr, Trp, Cys, Pro, Leu, Ile, Lys, and Arg residues will be avoided, Cys residues due to their potential for formation of unwanted disulfide bonds, Pro residues due to their hydrophobicity and lack of flexibility, and Lys and Arg residues due to risk of possible immunogenicity.

A convenient linker may be represented by the sequence of SEQ ID NO 2 represented in figure 2C.

This linker fragment is included in the loop in substitution of wildtype residues.

The deleted wildtype oligopeptide may have or not the same length than the linker oligopeptide.

In a specific embodiment, the deleted wildtype oligopeptide will be shorter than the fragment linker.

In another specific embodiment, the deleted wildtype will be longer than the fragment linker.

In still another specific embodiment, they will have the same length.

Wildtype oligopeptides that may be advantageously replaced by a linker in accordance with the present invention are represented by SEQ ID NO 15 and SEQ ID NO 16 (Figure 2D), and are corresponding respectively to sequences between positions 603 and 615, and between positions 558 and 622 of SEQ ID NO 1, and oligopeptides sequences between position 530 to 542 and between position 525 to 149 of SEQ ID NO 14.

According to a specific embodiment, the modified polypeptide in accordance with the present invention includes furthermore at least one mutation in its immunodominant region.

The immunodominant regions of which it is desired to modify the antigenicity, in accordance with the invention, may be chosen from those which give *in vitro* a cross-reaction, of the B type and/or the T type, with a host protein and in particular with IL-2.

In the appended Figure 2B, the peptide sequences of four regions of this region of gp41 555-577 (SEQ ID NO 3), 572-601 (SEQ ID NO 4), 590-620 (SEQ ID NO 5) and 628-663 (SEQ ID NO 6), are represented in which structural analogies and/or cross reactions were noted with IL-2. These regions are homologously found in the SEQ ID NO 14, and are respectively corresponding to peptide sequences found in positions 482-504, 499-529, 517-547, and 555-590.

Accordingly, within another aspect of the invention, there is provided a modified polypeptide as disclosed here-above and including furthermore at least a mutation in its immunodominant region to prevent a cross reaction of the B type and/or the T type with a host protein and in particular with human IL-2.

Are particularly interesting modified polypeptides according to the invention including furthermore at least a mutation in one of the immunodominant region represented by the peptide sequence SEQ ID NO 3, SEQ ID NO 4, SEQ ID NO 5 and/or SEQ ID NO 6.

In the present application, "mutation" refers to any modification of a region (optionally reduced to a single amino acid residue) of a polypeptide, by physical means, chemical means (covalent or noncovalent modification) and/or biological means (mutations by substitution, deletion and/or insertion of one or more amino acids), leading to the modification of the functional potentials of the constituent amino acid(s) of said region, termed "mutated region". By way of example, it is possible to carry out mutations leading to the abolition, acquisition and/or modulation of the properties of disulfide bridges, hydrogen bonds, electrostatic interactions and/or hydrophobic interactions, the modification of the capacity of a protein to form a heterocomplex, or alternatively, in the case of an oligomeric protein, the modification of the state of oligomerization or of the stability of the oligomer.

Some of the mutations decisive to impact this change in the antigenicity are disclosed in US 6 455 265 which teaching is hereby incorporated by reference herein in its entirety.

A modified polypeptide according to the instant invention is in particularly represented by the sequence SEQ ID NO 8 of Figure 1. This modified polypeptide has been derived from SEQ ID NO 14, wherein the oligopeptide sequence between positions 531 to 542 has been replaced with a linker corresponding to SEQ ID NO 2 (Figure 2C), and the oligopeptide sequence between positions 597 to 607 has been replaced by a His-Tag. In these sequences, an additional mutation has been carried-out in position 528 (numbering according to SEQ ID NO 14), wherein a trytophan residue has been replaced by an aspartate amino acid residue.

Others polypeptide sequences illustrating the present invention are represented by SEQ ID NO 17, SEQ ID NO 18, SEQ ID NO 19 and SEQ ID NO 20 (Figures 4 and 5).

According to another particular embodiment of the instant invention, the gp41 modified polypetide may also include modification such as truncation of a part of the amino acid sequence at the N or C-terminal extremities.

In an advantageously embodiment, the gp41 modified polypeptide is truncated at the N-terminal extremity.

Without wishing to be bound by any theory, the inventors hypothesize that such polypeptide modifications allow to some C-peptide sequences not being, for instance totally or partially, masked, in particular at the C-terminal extremity, which is therefore more readily accessible to antibodies.

More particularly, the gp41 modified polypeptide is a N-truncated oligopeptide. The length of the deletion is of a size ranging from 8 to 15 amino acid residues numbered from SEQ ID NO 1.

More advantageously, the gp41 modified polypeptide is truncated of at least 10, in particular of at least 12 amino acid residues at the N-terminal position.

An illustrative polypeptide sequence exhibiting N-terminal truncation and a replacement in the loop sequence of a oligopeptide sequence by a linker (SED ID NO 2) in accordance with the instant invention, is in particular represented by the sequence SEQ ID NO 21 (Figure 6).

According to another particular embodiment of the invention, it may be possible to include besides the loop modifications, carried-out to improve the solubility, and the modifications to improve the antigenicity of the gp41 polypeptide, additional modifications to improve solubility.

As particular embodiment of such modifications, mention may be made of amino acid residue mutation, as for example changing the tryptophan residue in position 528 (numbering SEQ ID NO 14) by a more hydrophilic amino acid residue, such as an aspartate.

Other similar mutations are illustrated by the gp41 modified polypeptides set forth in SEQ ID NO 17 and SEQ ID NO 19 wherein the tryptophan in position 124 in SEQ ID NO 17 and 130 in SEQ ID NO 19, that would correspond to the tryptophan in position 685 of SEQ ID NO 1 (not represented in this sequence), has been exchanged by an aspartate.

Such mutations are carried-out in the purpose of changing hydrophobic amino acid residues by more hydrophilic amino acid residues.

Other modifications to improve the solubility of the gp41 modified polypeptide may be deletion or insertion of amino acid residues.

As particular embodiment of such modifications, mention may be made of amino acid residues deletion, as for example deletion of amino acid residues tryptophan and leucine respectively in positions 122 and 123 of SEQ ID NO 17, and in positions 128 and 129 of SEQ ID NO 19 to give respectively the sequences set forth as SEQ ID NO 18 and SEQ ID NO 20.

Moreover, without departing of the ambit in the instant invention, the modified polypeptides in the accordance with the present invention may include additional modifications useful for laboratory experiments, such as fusion with a His-Tag or a fluorescent protein, as for example Green Fluorescent Protein.

Illustrative gp41 modified polypeptides according to the instant invention which carry additional modifications, as His-Tag, are set forth in SEQ ID NO 8, SEQ ID NO 17, SEQ ID NO 18, SEQ ID NO 19, SEQ ID NO 20 and, SEQ ID NO 21 (Figures 1, 4-6). The His-Tag has been added at the C-terminal extremity of those peptides.

To prepare the modified polypeptide according to the invention, it is possible to use any known methods of peptide synthesis or genetic engineering techniques, such as described in *Molecular cloning : a laboratory manual*, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

It is possible to isolate or to prepare a polynucleotide sequence encoding at least part of the envelopes gp41 of HIV-1 and, if desired, it is possible to introduce at this

stage, into the nucleotide sequence, the linker and optionally the mutation which make it possible to obtain a mutated product of translation which constitutes the modified polypeptide.

It is also possible to directly synthesize a modified polynucleotide sequence comprising one or more mutations encoding the modified polypeptide. The mutated polynucleotide sequences thus obtained are introduced in a known manner into an appropriate vector which makes it possible to express said polypeptide, optionally in modified form. Such a vector is for example *E. Coli*, a baculovirus or a mammalian cell. It is also possible to carry out the mutation on an unmodified polypeptide obtained according to one of the preceding methods.

Accordingly, the present invention is also directed to polynucleotide molecules, including DNA and RNA molecules, that encode the modified polypeptides disclosed above.

The polynucleotides of the present invention include both single-stranded and double-stranded molecules. A representative DNA sequence is set forth in SEQ ID NO 7 of Figure 1.

Additional DNA sequences encoding modified polypeptides can be readily generated by those of ordinary skill in the art based on the genetic code. Counterpart RNA sequences can be generated by substitution of U for T. Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among polynucleotide molecules encoding modified polypeptides.

The subject of the invention is also the use of a modified polypeptide, as defined above, in the preparation of a vaccine composition for preventing the pathogenic effects related to the infection of a host by a retrovirus HIV-1.

Thus, the modified polypeptide obtained according to the invention can serve as immunogenic agent in order to induce, by immunization, the formation of antibodies which can be used in particular in the treatment of retroviral infections, and the invention therefore also relates to the antibodies obtained in response to the immunization of animals (including humans), *in vivo* with the aid of the vaccine agent containing a modified polypeptide as above-described.

The pharmaceutical compositions containing such antibodies also constitute one of the subjects of the invention.

It is known from the person skilled in the art that there are various strains of a HIV viruses, and therefore various homologous forms of gp41 proteins.

It is also known from the person skilled in the art that knowing various homologous sequences of a given peptide or a protein, it is possible to design consensus sequences of such peptide or protein that display the sequence of amino acid residues the most frequently present in such protein or peptide.

Consequently, it is not going beyond the scope of the instant invention to carrying similar loop engineering and the other disclosed modifications by adapting them to a particular gp41 protein isoform to modify or to a consensus sequence to modify.

The following non limiting examples illustrate the invention.

Example 1: Construction of gp41-engineered loop of SEQ ID NO 8 by molecular biology

a) Design of the oligonucleotide primers

The gp41-engineered loop was constructed by PCR. The amplification of the N-helix and introduction of the hydrophilic linker was carried out by using the oligonucleotide primer gp41-NdeI (SEQ ID NO 9): 5' GG AAT CCA *CAT ATG CAG GCC AGA CAA TTA TTG* 3', and the oligonucleotide primer gp41-Bam1IL (SEQ ID NO 10): 5'ACC GTT GGA TCC ACC TCT ACC TCC ACT *GCT ACC* ***GTC*** *AAT CCC CAG GAG CTG TTG ATC* 3' (Figure 7).

These oligonucleotide primers were designed to respectively introduce the sites for restriction enzymes NdeI and BamHI (twice underlined into the oligonucleotides primers sequences). The sequences homologous to the gp41 gene into both oligonucleotides primers are written in italic. The oligonucleotide primer gp41-Bam1IL was also designed to introduce the oligopeptide linker SGGRGGS (SEQ ID NO 2) to replace the deleted portion of loop (corresponding to the once and twice underlined sequences) and a mutation to introduce in the position 528 (protein numbering SEQ ID NO 14) wherein a tryptophan has been replaced by an aspartate amino acid (bold triplet).

The amplification of the C-helix of gp41 protein was carried-out by PCR, using the oligonucleotide primer gp41-Bam2IL (SEQ ID NO 11): 5' GG AAT CCA GGA TCC *AAT GCT AGT TGG AGT AAT AAA TCT CTG GAA* 3', and the oligonucleotide primer gp41-XhoI (SEQ ID NO 12): 5' GCC CGG CTC GAG *ATC TAA TTC CAA TAA TTC TTG*

*TTC ATT CTT TTC* 3' (Figure 7). Those oligonucleotide primers were designed to respectively introduce the BamHI and the XhoI enzyme sites restrictions (sequences twice underlined). The sequences homologous to the gp41 gene are written in italic in both primers.

b) Conditions of PCR

The gp41 modified polypeptide was amplified from the gp41 matrix (SEQ ID NO 13) by PCR using the above-described oligonucleotide primers. Plasmid was used at 0.5 µg/µl, primers were used at 10 µM each, and dNTP were used at 10 mM each. The amplification was conducted using the DNA polymerase DyNazyme from Finnzymes. The amplification was initiated with a denaturing step of 5 minutes at 94 °C, following by 30 cycles, each comprising a one minute step at 94 °C (denaturing step), a one minute step at 60 °C (hybridization), and a one minute step at 72 °C (elongation), and the amplification was terminated by a last step of 10 minutes at 72 °C.

The purified PCR products were digested by NdeI-BamHI for the N-helix amplification and by BamHI and XhoI for the C-helix amplification. The two purified NdeI-BamHI and BamHI-XhoI fragments were ligated into the NdeI-XhoI sites of pET21b vector (Novagen®) resulting in pET21b-gp41-engineered loop.

The introduction of the ligated fragment into the XhoI site of the pET21b plasmid results in addition of a His-Tag at the C-terminus of the gp41 modified polypeptides preceded by a glutamate and leucine amino acid residues (see SEQ ID NO 8) pET21b-gp41-engineered loop products were transformed in DH5α.

The complete nucleotide sequence of gp41 modified polypeptide (SEQ ID NO 8) was determined by Genome Express (Grenoble). No mutation was detected.

Example 2: modified polypeptide reproduction in *E. coli*

a) Transformation

pET21b-gp41-engineered loop plasmid is transformed in expression *E. coli* strain (BL21(DE3)).

b) Expression tests

6 cultures of *E. coli* strain BL21(DE3) carrying the pET21b-gp41-engineered loop plasmid were grown at 37°C in Luria Broth until the optical density at 600 nm reached 0.6 (spectrophotometer Jasco V-530). The modified polypeptide was induced with 1 mM IPTG (isopropyl βD-thiogalactoside), and the culture continued for further 2 hours at 37°C. The gp41-engineered loop protein is expressed in E. coli as a 15 kDa protein specie.

Expression of proteins were controled by separation by SDS-4-12% PAGE and immunoblotting with antibodies anti-His tag.

c) Production

1) Culture

One liter of culture of BL21 (DE3)/pET21b-gp41-engineered loop was grown in Luria Broth at 37°C until the optical density at 600 nm reached the value of 6.0. The expression of gp41-engineered loop was induced by 1 mM IPTG, and the culture continued for a further 2 hours at 37°C. The culture was centrifuged (Centrifuge Beckman Coulter Avanti J20XP with rotor JLA 8-1000, 4000 x g, 30 min, 4°C) and the pellet was stored at -80°C.

2) Extraction of gp41 modified polypeptide

The pellet was resuspended with a sonication buffer (40 mL of Tris-HCl 50 mM pH8, NaCl 300 mM). Bacteria were disrupted by a 15 min sonication step on ice/ethanol (disintegrator UP200S amplitude 80%, coeff. 0.5) . Then the suspension was centrifuged at 40 000 x g during 30 min at 4°C to separate the soluble proteins (supernatant) from the insoluble proteins (pellet) (Centrifuge Beckman Coulter Avanti J20XP with rotor JA20).

The gp41 modified polypeptide is soluble in the sonication supernatant in majority (80%).

d) Purification of gp41 modified polypeptide

The sonication supernatant (# 50 mL) was filtered through a 0.2-µm filter. The presence of 6 His at the C-terminal extremity allows a purification by affinity chromatography columns.

1) Affinity chromatography

The affinity chromatography was conducted using an Akta FPLC (Fast-Pressure Liquid Chromatography, Amersham-Biosciences), and a Chelating Sepharose Fast Flow column (from Amersham-Pharmacia). The column was initially equilibrated with the passage of equilibration buffer A (Tris 50 mM pH8, NaCl 300 mM) (10 times the column volume).

Then 50 ml of sample containing gp41-engineered loop proteins were past through the column.

The contaminants were eluted in four steps with the passage of Buffer A comprising increasing amount of Buffer B (Tris 50 mM pH8, NaCl 300 mM, Imidazole 500 mM) in a volume corresponding to 10 times the column volume at each step (0, 20, 50 and 100 mM Imidazole). And finally the gp41 protein was eluted with the passage of 100 % of Buffer B, in a volume corresponding to ten times the column volume.

The flow rate was set up at 8 mL/min and the collected fraction volume was 2 mL.

The presence of proteins was detected using an UV lamp and measuring of the absorbance at 280 nm.

The first four fractions were collected and mixed (8 mL at 0.2 mg/mL).

2) Dialysis

The gp41-engineered loop protein (8 mL in the buffer B (Tris 50 mM pH8, NaCl 300 mM, Imidazole 500 mM) was dialyzed three times (two analysis of 1h30 and one overnight) at 4 °C against 300 mL of Tris 50 mM pH8, NaCl 200 mM, Imidazole 200 mM. The resulting sample was centrifuged during 30 min at 30 000 g, 4 °C.

From one liter of bacterial culture, 10 mL of purified gp41-engineered loop protein in Tris 50 mM pH8, NaCl 200 mM, Imidazole 200 mM at 0.2 mg/mL were obtained.

For the storage at -80 °C, glycerol 5% was added and the purified protein was frozen in liquid nitrogen. At this step, the protein concentration was about 0.2 mg/mL.

Example 3:

a) Determination of the oligomeric state

Gel filtration with a separation between 3 000 and 600 000 Dalton was performed to determine the oligomeric state of the gp41-engineered loop protein.

The determination of the oligomeric state was conducted using a Fast Performance Liquid Chromatography (System BioLogic, Bio-Rad), with a analytic chromatography column Superdex 200 HR 10/30 Amersham-pharmacia.

The column was equilibrated with 50 ml of Buffer A (Tris 50 mM pH8, NaCl 200 mM), then 250 ml of sample containing gp41-engineered loop protein (corresponding to 500 µg of gp41-engineered loop protein) were passed through the column.

The proteins were therefore eluted using 30 ml of buffer A at a flow rate of 0.25 mL/min. The collected volume fraction was 2 mL. The proteins were detected using UV lamp and measuring of the absorbance at 280 nm.

The resulting chromatogram (Figure 8) indicated that the gp41-engineered loop protein construct was produced in a soluble and stable trimeric form at 0.2 mg/mL.

The differences between the consensus sequence of the gp41 ectodomain and the gp41-engineered loop were of 13%. Among the 18 amino acids constituting the ectodomain, 6 were different and 12 were removed. The ectodomain of the gp41 protein represents 71% of the entire gp41 sequence.

Table I: Summary of results

| | |
|---|---|
| Construct | gp41 modified polypeptide |
| Expression system | *E. coli* |
| Yield per litre of culture | 10 mL of purified protein at 0.2 mg/mL |
| Expressed as | Soluble fraction |
| Purity after refolding | > 97% |
| pH behaviour | Soluble at pH superior or equal to 7 |
| Chromatography steps | Purification onto affinity chromatography |
| Oligomeric state | Trimeric soluble at 0.2 mg/mL |